



(12) 发明专利

(10) 授权公告号 CN 115862636 B

(45) 授权公告日 2023.07.07

(21) 申请号 202211450136.3

G10L 17/18 (2013.01)

(22) 申请日 2022.11.19

G10L 15/00 (2013.01)

(65) 同一申请的已公布的文献号

审查员 周小林

申请公布号 CN 115862636 A

(43) 申请公布日 2023.03.28

(73) 专利权人 杭州珍林网络技术有限公司

地址 310013 浙江省杭州市西湖区西溪银

座3幢1层138室

(72) 发明人 刘益平 孟令军 翁羚 张强

王卓辉 郑双喜 李海舰

(74) 专利代理机构 北京博识智信专利代理事务

所(普通合伙) 16067

专利代理师 徐佳慧

(51) Int. Cl.

G10L 17/04 (2013.01)

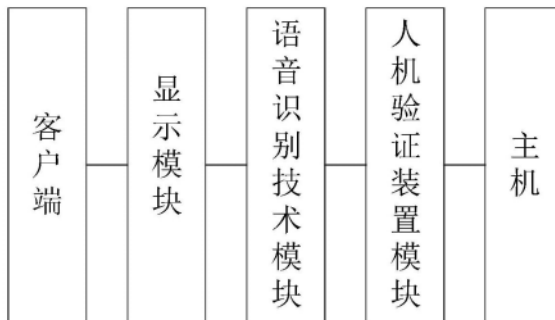
权利要求书3页 说明书9页 附图3页

(54) 发明名称

一种基于语音识别技术的互联网人机验证方法

(57) 摘要

本发明公开了一种基于语音识别技术的互联网人机验证方法,涉及语音识别技术、人机验证领域,包括显示模块、语音识别技术模块、人机验证装置模块,客户端的输出端与所述显示模块的输入端连接,所述显示模块的输出端与所述语音识别技术模块的输入端连接,所述语音识别技术模块的输出端与所述人机验证装置模块的输入端连接,所述人机验证装置模块的输出端与主机连接;通过简易的算法与语言识别技术,解决了人机验证不安全不便捷、训练时间长的问题,得到了安全有效的人机验证方法。本发明大大提高了人机验证的安全性及准确性。



1. 一种基于语音识别技术的互联网人机验证方法,其特征在于:包括:

显示模块,用于显示出客户端所需要录入的内容与人机验证规则;

语音识别技术模块,用于识别语音数据信息并将用户语音中的词汇内容转换为计算机可读的输入信息,以确认说话人讲话对象及讲话内容信息;

其中所述语音识别技术模块包括语音信号预处理模块、语音信号特征提取模块、训练识别网络模块和语音信号比对模块;所述语音信号预处理模块的输入端接收用户语音输出信息,所述语音信号预处理模块的输出端连接所述语音信号特征提取模块的输入端,所述语音信号特征提取模块的输出端连接所述训练识别网络模块的输入端,所述训练识别网络模块的输出端连接所述语音信号比对模块的输入端,所述语音信号比对模块的输入端连接主机;其中语音信号预处理模块用于剔除语音数据信息中的粗大信息,语音信号特征提取模块用于提取语音的频域和时域数据信息,训练识别网络模块通过改进型DBN算法模型对提取到的语音数据信息进行学习、训练和识别输出,所述语音信号比对模块用于将输入的数据信息和标准数据信息进行对比;人机验证装置模块,用于对录入的语音信号进行验证,区分客户端是否为用户或机器;并对录入的语音信号进行故障诊断,以提高语音数据信息故障诊断能力;其中,

所述人机验证装置包括显示模块、获取模块、调整模块和验证模块;

所述显示模块用于显示人机验证装置的界面,并显示客户端发送的身份验证请求;

所述获取模块用于获取客户端发出的由语音识别技术模块输出的语音信号,并基于预先存储的客户端语音信号识别所录入的语音信号是否正确;

所述调整模块设置有音频信息时域和频域转换模块,用于通过转换音频数据信息,以满足音频数据信息的实时、动态信息转换,提高动态语音识别能力和互联网人机交互和计算能力;

所述验证模块用于基于语音信号特征,验证客户端为真用户或机器,并输出身份验证通过或失败的身份验证结果;

所述显示模块的输出端连接获取模块的输入端,所述获取模块的输出端连接调整模块的输入端,所述调整模块的输出端连接验证模块的输入端;语音识别方法模型,用于对录入的声音信号进行识别,以提高不同场景下语音数据信息识别能力;

其中,客户端的输出端与所述显示模块的输入端连接,所述显示模块的输出端与所述语音识别技术模块的输入端连接,所述语音识别技术模块的输出端与所述人机验证装置模块的输入端连接,所述人机验证装置模块的输出端与主机连接;

人机验证装置基于客户端发起的验证请求,用户端录入语音,人机验证装置获取用户端发出的由语音识别技术模块输出的语音信号,并基于预先存储的客户端语音信号,识别所录入的语音信号是否正确,基于语音信号特征验证客户端为真人或机器,并输出身份验证通过或失败的身份验证结果;

语音信号预处理模块包括预滤波模块、预加重模块、短时加窗处理模块和端点监测模块;语音信号输入至所述预滤波模块的输入端,所述预滤波模块的输出端连接所述预加重模块的输入端,所述预加重模块的输出端连接所述短时加窗处理模块的输入端,所述短时加窗处理模块的输出端连接所述端点监测模块的输入端;

所述预滤波模块为带通滤波器,下截止频率为100Hz,上截止频率为3400Hz;所述预加

重模块用于将语音信号从低频升到高频收集不同的信噪比,所述预加重模块使用一阶FIR滤波器对语音数据信息进行滤波输出函数为:

$$y(n) = x(n) - ax(n-1), 0.9 \leq a \leq 0.97 \quad (1)$$

在公式(1)中,a值取0.9386,y(n)表示一阶FIR滤波器的输出,x(n)表示一阶FIR滤波器的输入,x(n-1)表示一阶FIR滤波器上一时刻的输入;

所述短时加窗处理模块用于对语音信号进行加窗处理,通过窗序列{w(m)}截取一段语音信号并对输出的语音信息进行以下输出为:

$$Q_n = \sum_{m=-\infty}^{\infty} T[x(m)] \cdot w(n-m) \quad (2)$$

在公式(2)中,T[]表示运算函数,{x(m)}表示输入信号序列,时间依赖处理方法是:

当T[x(m)]为 $x^2(m)$ 时, Q_n 相应于短时能量;

当T[x(m)]为 $\text{sgn}[x(m)] - \text{sgn}[x(m-1)]$ 时, Q_n 为短时平均过零率;

当T[x(m)]为 $x(m) * x(m+k)$ 时, Q_n 相应于短时自相关函数;

分帧均采用汉明窗,帧长为256点,帧移96点,汉明窗窗函数为:

$$w(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n/(L-1)), & 0 \leq n \leq L-1 \\ 0, & \text{其他} \end{cases} \quad (3)$$

式(3)中,L为窗长,汉明窗的主瓣宽度b为1.19Hz,旁瓣高度a1为-43dB;

所述端点监测模块用于找出语音信号的起止点,正确找出语音信号的开始和终止使系统仅处理有效语音信号;端点监测模块的工作方法:

步骤一:计算出输入语音信号音节的最大短时平均幅度 M_{\max} ,语音信号{x(n)}的短时平均幅度的公式函数为:

$$M_n = \sum_{m=-\infty}^{\infty} |x(m)| \cdot w(n-m) \quad (4)$$

由公式(4)计算出该音节的最大短时平均幅度 M_{\max} ;

步骤二:设置判定条件:如果出现语音的短时平均幅度降低到最大短时平均幅度的1/16,即 $M_n < (1/16)M_{\max}$ 时,为语音信号结束点;

所述DBN算法模型包括模板匹配法模块、隐马尔科夫模型和人工神经网络模块;其中所述模板匹配法模块的输出端与隐马尔科夫模型的输入端连接,隐马尔科夫模型的输出端与人工神经网络模块的输入端连接;

所述模板匹配法模块,把时间规整和距离测度计算结合,将模板库中录入语音的特征矢量序列依次与模板库中的每个模板进行相似度比较,将相似度高者作为识别结果输出;

所述隐马尔科夫模型通过对语音数据信息进行向量序列观察,将所观测到的语音数据信息通过概率密度分布表现为各种状态,立体直观对语音数据向量观测;其中所述隐马尔科夫模型工作过程包含状态转移序列和每次转移时输出符号组成的观察序列;

所述人工神经网络模块包括4个RBM神经网络模块和2个BP神经网络模块;

隐马尔科夫模型通过动态前向算法实现语音数据信息的处理;

人工神经网络模块的工作方法为:

步骤一、设置RBM神经网络模块和BP神经网络模块,不同RBM神经网络模块彼此首尾连

接,并在第三个RBM神经网络模块输出端连接BP神经网络模块,所述BP神经网络模块的输出端与第四个RBM神经网络模块输出端连接;

步骤二、设置RBM神经网络模块和BP神经网络模块的权重和偏置参数;RBM神经网络模块中的RBM由一层可见层和一层隐藏层组成,在第一个RBM神经网络模块输入端设置权重和偏置参数,偏置参数用a、b表示,w表示连接权重,v表示输入量,h表示输出量;经过学习训练得到a、b和w;RBM神经网络模块输出函数为:

$$E(v,h) = -a^T v - b^T h - h^T W v \quad (5)$$

在公式(5)中,W表示权重函数输出量,E(v,h)表示RBM神经网络模块输出函数,T表示RBM神经网络模块计算时间,不断输入的音频输入输出量之间的概率密度分布函数表达式为:

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)} \quad (6)$$

式(6)中,z表示分配函数常数,分配函数表示在所有节点下 $e^{-E(v,h)}$ 的和,为了确保概率密度分布函数的积分为1;可见层节点激活状态的情况下,隐藏层的是否激活的条件概率方程式为:

$$P(v|h) = \prod_{j=1}^n P(h_j|v) \quad (7)$$

单个隐藏节点激活的概率表达式为:

$$P(h_j = 1|v) = \sigma \left(b_i + \sum_{j=1}^n w_{ij} v_i \right) \quad (8)$$

式(8)中, σ 表示sigmoid函数;

通过公式(5) - (8)训练集,最终训练DBN算法模型中的偏置和连接权重,最终输出DBN算法模型所输出语音信息的频率预测模型;

所述获取模块设置有音频信息转换接口;

验证模块输出的表达式为:

$$\begin{cases} Error = f(x_i) - g(x_i) \\ MRE = \frac{1}{N} \sum_{i=1}^N \frac{|f(x_i) - g(x_i)|}{|f(x_i)|} \\ MAE = \max(|f(x_i) - g(x_i)|) \\ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(x_i) - g(x_i))^2} \end{cases} \quad (9)$$

在公式(9)中,Error表示绝对误差值,MRE表示平均相对误差,MAE表示最大绝对误差,RMSE表示均方误差,其中i表示所验证语音数据信息样本编号,N表示输入语音数据信息的样本总数, $f(x_i)$ 和 $g(x_i)$ 分别表示合成输入语音样本的值和数据库语音数据样本的值。

一种基于语音识别技术的互联网人机验证方法

技术领域

[0001] 本发明涉及音频数据信息处理技术领域,更具体地说,尤其涉及一种语音识别技术的互联网人机验证方法。

背景技术

[0002] 人机验证技术是一种全自动区分计算机和用户的图灵测试技术,是一种区分用户是计算机还是人的公共全自动程序。人机验证技术的诞生是伴随着计算机技的发展和黑客技术的不断提高而出现的,黑客利用自动注册工具批量注册网站账号和灌水;随着科技的发展,语音识别技术的研究不断深入,人机验证技术以前所使用的字符型图片验证码验证和行为式验证码验证,已经不足以很好实现人机验证,无法做到安全有效进行人机验证,在面对多种数据信息处理时,用户难以对输入的音频数据信息分解、处理和计算,难以实现语音识别,进而降低了互联网人机验证能力,因此如何提高音频数据信息的处理和互联网人机交互能力就是亟待解决的技术问题。

发明内容

[0003] 针对上述技术的不足,本发明公开一种基于语音识别技术的互联网人机验证方法,通过语音识别技术,实现互联网人机验证,提高了人机验证的安全性和有效性。通过构建语音识别算法模型,在计算过程中通过语音数据信息分解、处理和计算,提高了音频数据信息的应用、识别能力。

[0004] 为了实现上述技术效果,本发明采用以下技术方案:

[0005] 一种基于语音识别技术的互联网人机验证方法,其中包括:

[0006] 显示模块,用于显示出客户端所需要录入的内容与人机验证规则;

[0007] 语音识别技术模块,用于识别语音数据信息并将将用户语音中的词汇内容转换为计算机可读的输入信息,以确认说话人讲话对象及讲话内容信息;

[0008] 其中所述语音识别技术模块包括语音信号预处理模块、语音信号特征提取模块、训练识别网络模块和语言信号比对模块;所述语言信号预处理模块的输入端接收用户语音输出信息,所述语言信号预处理模块的输出端连接所述语言信号特征提取模块的输入端,所述语言信号特征提取模块的输出端连接所述训练识别网络模块的输入端,所述训练识别网络模块的输出端连接所述语言信号比对模块的输入端,所述语言信号比对模块的输入端连接主机;其中语音信号预处理模块用于剔除语音数据信息中的粗大信息,语音信号特征提取模块用于提取语音的频域和时域数据信息,训练识别网络模块通过改进型DBN算法模型对提取到的语音数据信息进行学习、训练和识别输出,所述语言信号比对模块用于将输入的数据信息和标准数据信息进行对比;人机验证装置模块,用于对录入的语音信号进行验证,区分客户端是否为用户或机器;并对录入的语音信号进行故障诊断,以提高语音数据信息故障诊断能力;其中所述人机验证装置包括显示模块、获取模块、调整模块和验证模块;所述显示模块用于显示人机验证装置的界面,并显示客户端发送的身份验证请求;所述

获取模块用于获取用户端发出的由语音识别模块输出的语音信号,并基于预先存储的客户端语言信号识别所录入的语音信号是否正确;所述确定模块用于若所述录入的语音信号正确,基于语音信号特征判断确定客户端为真用户或机器;所述验证模块用于基于语音信号特征,验证客户端为真用户或机器,并输出身份验证通过或失败的身份验证结果;所述显示模块的输出端连接获取模块的输入端,所述获取模块的输出端连接调整模块的输入端,所述调整模块的输出端连接验证模块的输入端;语音识别方法模型,用于对录入的声音信号进行识别,以提高不同场景下语音数据信息识别能力;

[0009] 其中,客户端的输出端与所述显示模块的输入端连接,所述显示模块的输出端与所述语音识别技术模块的输入端连接,所述语音识别技术模块的输出端与所述人机验证装置模块的输入端连接,所述人机验证装置模块的输出端与主机连接。

[0010] 作为本发明进一步的技术方案,语音信号预处理模块包括预滤波模块、预加重模块、短时加窗处理模块和端点监测模块;语音信号输入至所述预滤波模块的输入端,所述预滤波模块的输出端连接所述预加重模块的输入端,所述预加重模块的输出端连接所述短时加窗处理模块的输入端,所述短时加窗处理模块的输出端连接所述端点监测模块的输入端。

[0011] 作为本发明进一步的技术方案,所述预滤波模块为带通滤波器,下截止频率为100Hz,上截止频率为3400Hz;所述预加重模块用于将语音信号从低频升到高频收集不同的信噪比,所述预加重模块使用一阶FIR滤波器对语音数据信息进行滤波输出函数为:

$$[0012] \quad y(n) = x(n) - ax(n-1), 0.9 \leq a \leq 0.97 \quad (1)$$

[0013] 在公式(1)中, a 值取0.9386, $y(n)$ 表示一阶FIR滤波器的输出, $x(n)$ 表示一阶FIR滤波器的输入, $x(n-1)$ 表示一阶FIR滤波器上一时刻的输入;

[0014] 所述短时加窗处理模块用于对语音信号进行加窗处理,通过窗序列 $\{w(m)\}$ 截取一段语音信号并对输出的语音信息进行以下输出为:

$$[0015] \quad Q_n = \sum_{m=-\infty}^{\infty} T[x(m)] \cdot w(n-m) \quad (2)$$

[0016] 在公式(2)中,式(2)中 $T[\]$ 表示运算函数, $\{x(m)\}$ 表示输入信号序列,时间依赖处理方法是:

[0017] 当 $T[x(m)]$ 为 $x^2(m)$ 时, Q_n 相应于短时能量;

[0018] 当 $T[x(m)]$ 为 $\text{sgn}[x(m)] - \text{sgn}[x(mI)]$ 时, Q_n 为短时平均过零率;

[0019] 当 $T[x(m)]$ 为 $x(m) * x(m+k)$ 时, Q_n 相应于短时自相关函数;

[0020] 分帧均采用汉明窗,帧长为256点(32ms),帧移96点(12ms),汉明窗窗函数为:

$$[0021] \quad w(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n / (L-1)), & 0 \leq n \leq L-1 \\ 0, & \text{其他} \end{cases} \quad (3)$$

[0022] 式(3)中, L 为窗长,汉明窗的主瓣宽度 b 为1.19Hz,旁瓣高度 a_1 为-43dB。

[0023] 作为本发明进一步的技术方案,所述端点监测模块用于找出语音信号的起止点,正确找出语音信号的开始和终止使系统仅处理有效语音信号;端点监测模块的工作方法:

[0024] 步骤一:计算出输入语音信号音节的最大短时平均幅度 M_{\max} ,语音信号 $\{x(n)\}$ 的短

时平均幅度的公式函数为:

$$[0025] \quad M_n = \sum_{m=-\infty}^{\infty} |x(m)| \cdot w(n-m) \quad (4)$$

[0026] 由公式(4)计算出该音节的最大短时平均幅度 M_{\max} ;

[0027] 步骤二:设置判定条件:如果出现语音的短时平均幅度降低到最大短时平均幅度的1/16,即 $M_n < (1/16)M_{\max}$ 时,为语音信号结束点。

[0028] 作为本发明进一步的技术方案,所述DBN算法模型包括模板匹配法模块、隐马尔科夫模型和人工神经网络模块;其中所述模板匹配法模块的输出端与隐马尔科夫模型的输入端连接,隐马尔科夫模型的输出端与人工神经网络模块的输入端连接;

[0029] 所述模板匹配模块把时间规整和距离测度计算结合,将模板库中录入语音的特征矢量序列依次与模板库中的每个模板进行相似度比较,将相似度高者作为识别结果输出;

[0030] 所述隐马尔科夫模型通过对语音数据信息进行向量序列观察,将所观测到的语音数据信息通过概率密度分布表现为各种状态,立体直观对语音数据向量观测;其中所述隐马尔科夫模型工作过程包含状态转移序列和每次转移时输出符号组成的观察序列;

[0031] 所述人工神经网络模块包括4个RBM神经网络模块和2个BP神经网络模块。

[0032] 作为本发明进一步的技术方案,隐马尔科夫模型通过动态前向算法实现语音数据信息的处理。

[0033] 作为本发明进一步的技术方案,人工神经网络模块的工作方法为:

[0034] 步骤一、设置RBM神经网络模块和BP神经网络模块,不同RBM神经网络模块彼此首位连接,并在第三个RBM神经网络模块输出端连接BP神经网络模块,所述BP神经网络模块的输出端与第四个RBM神经网络模块输出端连接;

[0035] 步骤二、设置RBM神经网络模块和BP神经网络模块的权重和偏置参数;RBM神经网络模块中的RBM由一层可见层和一层隐藏层组成,在第一个RBM神经网络模块输入端设置权重和偏置参数,偏置参数用 a 、 b 表示, w 表示连接权重, v 表示输入量, h 表示输出量;经过学习训练可以得到 a 、 b 和 w ;RBM神经网络模块输出函数为:

$$[0036] \quad E(v, h) = -a^T v - b^T h - h^T W v \quad (5)$$

[0037] 在公式(5)中, w 表示权重函数输出量, $E(v, h)$ 表示RBM神经网络模块输出函数, T 表示RBM神经网络模块计算时间,不断输入的音频输入输出量之间的概率密度分布函数表达式为:

$$[0038] \quad P(v, h) = \frac{1}{Z} e^{-E(v, h)} \quad (6)$$

[0039] 式(6)中, z 表示分配函数常数,分配函数表示在所有节点下 $e^{-E(v, h)}$ 的和,为了确保概率密度分布函数的积分为1;可见层节点激活状态的情况下,隐藏层的是否激活的条件概率方程式为:

$$[0040] \quad P(v|h) = \prod_{j=1}^n P(h_j|v) \quad (7)$$

[0041] 单个隐藏节点激活的概率表达式为:

$$[0042] \quad P(h_j = 1|v) = \sigma \left(b_j + \sum_{i=1}^n w_{ij} v_i \right) \quad (8)$$

[0043] 式(8)中, σ 表示sigmoid函数。

[0044] 通过公式(5)-(8)训练集,最终训练DBN算法模型中的偏置和连接权重,最终输出DBN算法模型所输出语音信息的频率预测模型。

[0045] 作为本发明进一步的技术方案,所述获取模块设置有音频信息转换接口。

[0046] 作为本发明进一步的技术方案,所述调整模块设置有音频信息时阈和频域转换模块。

[0047] 作为本发明进一步的技术方案,验证模块输出的表达式为:

$$[0048] \quad \begin{cases} Error = f(x_i) - g(x_i) \\ MRE = \frac{1}{N} \sum_{i=1}^N \frac{|f(x_i) - g(x_i)|}{|f(x_i)|} \\ MAE = \max(|f(x_i) - g(x_i)|) \\ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(x_i) - g(x_i))^2} \end{cases} \quad (9)$$

[0049] 在公式(9)中, $Error$ 表示绝对误差值, MRE 表示平均相对误差, MAE 表示最大绝对误差, $RMSE$ 表示均方误差, 其中 i 表示所验证语音数据信息样本编号, N 表示输入语音数据信息的样本总数, $f(x_i)$ 和 $g(x_i)$ 分别表示合成输入语音样本的值和数据库语音数据样本的值。

[0050] 积极有益效果

[0051] 本发明的技术效果和优点:本发明提供了一种基于语音识别技术的互联网人机验证方法,与现有技术相比,本发明通过简易的算法与语言识别技术,解决了人机验证不安全不便捷,训练时间长等众多方面的复杂问题,得到了安全准确的人机验证方法。

[0052] 本发明通过构建具有显示模块、语音识别技术模块和验证模块,提高了语音信息的识别能力,所述语音识别技术模块包括语音信号预处理模块、语音信号特征提取模块、训练识别网络模块和语言信号比对模块;语音信号预处理模块用于剔除语音数据信息中的粗大信息,语音信号特征提取模块用于提取语音的频域和时域数据信息,训练识别网络模块通过改进型DBN算法模型对提取到的语音数据信息进行学习、训练和识别输出,所述语言信号比对模块用于将输入的数据信息和标准数据信息进行对比;人机验证装置模块,用于对录入的语音信号进行验证,通过对录入的声音信号进行识别,以提高不同场景下语音数据信息识别能力,从而提高了本发明的应用能力。

附图说明

[0053] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图,其中:

- [0054] 图1为本发明一种基于语音识别技术的互联网人机验证方法的系统框图；
- [0055] 图2为本发明中语言识别技术系统框图；
- [0056] 图3为本发明中语音信号预处理系统框图；
- [0057] 图4为本发明中人机验证装置系统框图。

具体实施方式

[0058] 以下结合附图对本发明的优选实施例进行说明,应当理解,此处所描述的实施例仅用于说明和解释本发明,并不用于限定本发明。

[0059] 如图1-图4所示,一种基于语音识别技术的互联网人机验证方法,其中包括:

[0060] 显示模块,用于显示出客户端所需要录入的内容与人机验证规则;

[0061] 语音识别技术模块,用于识别语音数据信息并将用户语音中的词汇内容转换为计算机可读的输入信息,以确认说话人讲话对象及讲话内容信息;

[0062] 其中所述语音识别技术模块包括语音信号预处理模块、语音信号特征提取模块、训练识别网络模块和语言信号比对模块;所述语言信号预处理模块的输入端接收用户语音输出信息,所述语言信号预处理模块的输出端连接所述语言信号特征提取模块的输入端,所述语言信号特征提取模块的输出端连接所述训练识别网络模块的输入端,所述训练识别网络模块的输出端连接所述语言信号比对模块的输入端,所述语言信号比对模块的输入端连接主机;其中语音信号预处理模块用于剔除语音数据信息中的粗大信息,语音信号特征提取模块用于提取语音的频域和时域数据信息,训练识别网络模块通过改进型DBN算法模型对提取到的语音数据信息进行学习、训练和识别输出,所述语言信号比对模块用于将输入的数据信息和标准数据信息进行对比;人机验证装置模块,用于对录入的语音信号进行验证,区分客户端是否为用户或机器;并对录入的语音信号进行故障诊断,以提高语音数据信息故障诊断能力;其中所述人机验证装置包括显示模块、获取模块、调整模块和验证模块;所述显示模块用于显示人机验证装置的界面,并显示客户端发送的身份验证请求;所述获取模块用于获取客户端发出的由语音识别模块输出的语音信号,并基于预先存储的客户端语言信号识别所录入的语音信号是否正确;所述确定模块用于若所述录入的语音信号正确,基于语音信号特征判断确定客户端为真用户或机器;所述验证模块用于基于语音信号特征,验证客户端为真用户或机器,并输出身份验证通过或失败的身份验证结果;所述显示模块的输出端连接获取模块的输入端,所述获取模块的输出端连接调整模块的输入端,所述调整模块的输出端连接验证模块的输入端;语音识别方法模型,用于对录入的声音信号进行识别,以提高不同场景下语音数据信息识别能力;

[0063] 其中,客户端的输出端与所述显示模块的输入端连接,所述显示模块的输出端与所述语音识别技术模块的输入端连接,所述语音识别技术模块的输出端与所述人机验证装置模块的输入端连接,所述人机验证装置模块的输出端与主机连接。

[0064] 在上述实施例中,语音识别技术模块在训练阶段,语音信号经过预处理、特征提取后,得到代表此语音的特征矢量,用此特征矢量去训练声学模型,多个语音的声学模型就构成了语音的参考模式库。在识别阶段,输入语音信号经过预处理、特征提取后,用此特征矢量与参考模式库里的每个模型进行比较,按照一定的准则判决出识别结果。

[0065] 在本发明中,语音信号预处理模块包括预滤波模块、预加重模块、短时加窗处理模

块和端点监测模块；语音信号输入至所述预滤波模块的输入端，所述预滤波模块的输出端连接所述预加重模块的输入端，所述预加重模块的输出端连接所述短时加窗处理模块的输入端，所述短时加窗处理模块的输出端连接所述端点监测模块的输入端。

[0066] 在具体实施例中，人机验证装置基于客户端发起的验证请求，用户端录入语音，人机验证装置获取用户端发出的由语音识别模块输出的语音信号，并基于预先存储的客户端语音信号，识别所录入的语音信号是否正确，基于语音信号特征验证客户端为真人或机器，并输出身份验证通过或失败的身份验证结果。

[0067] 在具体实施例中，所述预滤波模块为带通滤波器，下截止频率为100Hz，上截止频率为3400Hz；所述预加重模块用于将语音信号从低频升到高频收集不同的信噪比。预滤波又称之为反混叠滤波，是指滤除高于1/2 采样频率的信号成分或噪声，使信号带宽限制在某个范围内；否则，如果采样率不满足采样定理，则会产生频谱混叠，此时信号中的高频成分将产生失真；而工频干扰指50Hz的电源干扰。

[0068] 由于语音信号的功率受口鼻的影响，800Hz以上的频段有6dB的衰减，即，求频谱时，频率越高则信号幅度越小。这常常使得频谱分析时，从低频到高频的整个频带内不能使用相同的信噪比；所述预加重模块用于将语音信号从低频升到高频收集不同的信噪比，所述预加重模块使用一阶FIR滤波器对语音数据信息进行滤波输出函数为：

$$[0069] \quad y(n) = x(n) - ax(n-1), 0.9 \leq a \leq 0.97 \quad (1)$$

[0070] 在公式(1)中， a 值取0.9386， $y(n)$ 表示一阶FIR滤波器的输出， $x(n)$ 表示一阶FIR滤波器的输入， $x(n-1)$ 表示一阶FIR滤波器上一时刻的输入；

[0071] 所述短时加窗处理模块用于对语音信号进行加窗处理，通过窗序列 $\{w(m)\}$ 截取一段语音信号并对输出的语音信息进行以下输出为：

$$[0072] \quad Q_n = \sum_{m=-\infty}^{\infty} T[x(m)] \cdot w(n-m) \quad (2)$$

[0073] 在公式(2)中，式(2)中 $T[\]$ 表示运算函数， $\{x(m)\}$ 表示输入信号序列。在具体实施例中。语音通常在 10~30ms之内，其特性基本保持不变，即在短时间内相对稳定。基于这样的考虑，对语音信号的时域分析和频域处理必须建立在“短时分析”的基础上，即将语音信号分成一段一段地分析，每一段定义为一“帧”，帧长一般取10~30ms，因此所述短时加窗处理模块用于对语音信号进行加窗处理，为了得到短时的语音信号，所述短时加窗处理模块用一个长度有限的窗序列 $\{w(m)\}$ 截取一段语音信号分析。

[0074] 时间依赖处理方法是：

[0075] 当 $T[x(m)]$ 为 $x^2(m)$ 时， Q_n 相应于短时能量；

[0076] 当 $T[x(m)]$ 为 $\text{sgn}[x(m)] - \text{sgn}[x(mI)]$ 时， Q_n 为短时平均过零率；

[0077] 当 $T[x(m)]$ 为 $x(m) * x(m+k)$ 时， Q_n 相应于短时自相关函数；

[0078] 分帧均采用汉明窗，帧长为256点(32ms)，帧移96点(12ms)，汉明窗窗函数为：

$$[0079] \quad w(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n/(L-1)), & 0 \leq n \leq L-1 \\ 0, & \text{其他} \end{cases} \quad (3)$$

[0080] 式(3)中， L 为窗长，汉明窗的主瓣宽度 b 为1.19Hz，旁瓣高度 a_1 为-43dB。

[0081] 在本发明具体实施例中,所述端点监测模块用于找出语音信号的起止点,正确找出语音信号的开始和终止使系统仅处理有效语音信号;端点监测模块的工作方法:

[0082] 步骤一:计算出输入语音信号音节的最大短时平均幅度 M_{\max} ,语音信号 $\{x(n)\}$ 的短时平均幅度的公式函数为:

$$[0083] \quad M_n = \sum_{m=-\infty}^{\infty} |x(m)| \cdot w(n-m) \quad (4)$$

[0084] 由公式(4)计算出该音节的最大短时平均幅度 M_{\max} ;

[0085] 步骤二:设置判定条件:如果出现语音的短时平均幅度降低到最大短时平均幅度的1/16,即 $M_n < (1/16)M_{\max}$ 时,为语音信号结束点。

[0086] 在本发明具体实施例中,所述DBN算法模型包括模板匹配法模块、隐马尔科夫模型和人工神经网络模块;其中所述模板匹配法模块的输出端与隐马尔科夫模型的输入端连接,隐马尔科夫模型的输出端与人工神经网络模块的输入端连接;

[0087] 所述模板匹配模块把时间规整和距离测度计算结合,将模板库中录入语音的特征矢量序列依次与模板库中的每个模板进行相似度比较,将相似度高者作为识别结果输出;

[0088] 所述隐马尔科夫模型通过对语音数据信息进行向量序列观察,将所观测到的语音数据信息通过概率密度分布表现为各种状态,立体直观对语音数据向量观测;其中所述隐马尔科夫模型工作过程包含状态转移序列和每次转移时输出符号组成的观察序列;

[0089] 所述人工神经网络模块包括4个RBM神经网络模块和2个BP神经网络模块。在具体实施例中,这种方法能力提高人工神经网络模块的计算能力和训练速度。给定隐马尔科夫模型,也就是在模型参数已知的情况下,想找到观察序列的概率。在本实施例中,可以描述为与语音识别有关系的内容,通过观察序列以及隐藏的状态,对互联网人机进行验证。

[0090] 在本发明具体实施例中,隐马尔科夫模型通过动态前向算法实现语音数据信息的处理。网格中的每一列都显示了可能语音状态,并且每一列中的每个状态都与相邻列中的每一个状态相连。而其状态间的转移都由状态转移矩阵提供一个概率。在每一列下面都是某个时间点上的观察状态,给定任一个隐藏状态所得到的观察状态的概率由混淆矩阵提供这种方法能够提高隐马尔科夫模型的工作效率和精度。

[0091] 在本发明具体实施例中,人工神经网络模块的工作方法为:

[0092] 步骤一、设置RBM神经网络模块和BP神经网络模块,不同RBM神经网络模块彼此首位连接,并在第三个RBM神经网络模块输出端连接BP神经网络模块,所述BP神经网络模块的输出端与第四个RBM神经网络模块输出端连接;在具体实施例中,通过将二者结合起来,以提高语音信息识别能力。

[0093] 步骤二、设置RBM神经网络模块和BP神经网络模块的权重和偏置参数;RBM神经网络模块中的RBM由一层可见层和一层隐藏层组成,在第一个RBM神经网络模块输入端设置权重和偏置参数,偏置参数用 a 、 b 表示, w 表示连接权重, v 表示输入量, h 表示输出量;经过学习训练可以得到 a 、 b 和 w ;RBM神经网络模块输出函数为:

$$[0094] \quad E(v, h) = -a^T v - b^T h - h^T W v \quad (5)$$

[0095] 在公式(5)中, w 表示权重函数输出量, $E(v, h)$ 表示RBM神经网络模块输出函数, T

表示RBM神经网络模块计算时间,不断输入的音频输入输出量之间的概率密度分布函数表达式为:

$$[0096] \quad P(v, h) = \frac{1}{Z} e^{-E(v, h)} \quad (6)$$

[0097] 式(6)中, z 表示分配函数常数,分配函数表示在所有节点下 $e^{-E(v, h)}$ 的和,为了确保概率密度分布函数的积分为1;可见层节点激活状态的情况下,隐藏层的是否激活的条件概率方程式为:

$$[0098] \quad P(v|h) = \prod_{j=1}^n P(h_j | v) \quad (7)$$

[0099] 单个隐藏节点激活的概率表达式为:

$$[0100] \quad P(h_j = 1 | v) = \sigma \left(b_i + \sum_{j=1}^n w_{ij} v_i \right) \quad (8)$$

[0101] 式(8)中, σ 表示sigmoid函数。

[0102] 通过公式(5)-(8)训练集,最终训练DBN算法模型中的偏置和连接权重,最终输出DBN算法模型所输出语音信息的频率预测模型。在具体实施例中,通过设置DBN算法模型中的偏置和连接权重,将反映人工神经网络模块不同参数和数据信息的值,一步一步构建出来,进而提高人工神经网络模块工作能力。

[0103] 在本发明具体实施例中,所述获取模块设置有音频信息转换接口。这种方法便于接收语音,比如音频接口,蓝牙模块、无线通信接口,支撑音频信息的接口等。

[0104] 在本发明具体实施例中,所述调整模块设置有音频信息时阈和频域转换模块,比如通过转换音频数据信息,以满足音频数据信息的实时、动态信息转换,提高动态语音识别能力和互联网人机交互和计算能力。

[0105] 在本发明具体实施例中,验证模块输出的表达式为:

$$[0106] \quad \begin{cases} Error = f(x_i) - g(x_i) \\ MRE = \frac{1}{N} \sum_{i=1}^N \frac{|f(x_i) - g(x_i)|}{|f(x_i)|} \\ MAE = \max(|f(x_i) - g(x_i)|) \\ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(x_i) - g(x_i))^2} \end{cases} \quad (9)$$

[0107] 在公式(9)中, $Error$ 表示绝对误差值, MRE 表示平均相对误差, MAE 表示最大绝对误差, $RMSE$ 表示均方误差,其中 i 表示所验证语音数据信息样本编号, N 表示输入语音数据信息的样本总数, $f(x_i)$ 和 $g(x_i)$ 分别表示合成输入语音样本的值和数据库语音数据样本的值。在具体实施例中,采用不同验证公式,验证本发明技术效果。

[0108] 虽然以上描述了本发明的具体实施方式,但是本领域的技术人员应当理解,这些

具体实施方式仅是举例说明,本领域的技术人员在不脱离本发明的原理和实质的情况下,可以对上述方法和系统的细节进行各种省略、替换和改变;例如,合并上述方法步骤,从而按照实质相同的方法执行实质相同的功能以实现实质相同的结果则属于本发明的范围;因此,本发明的范围仅由所附权利要求书限定。

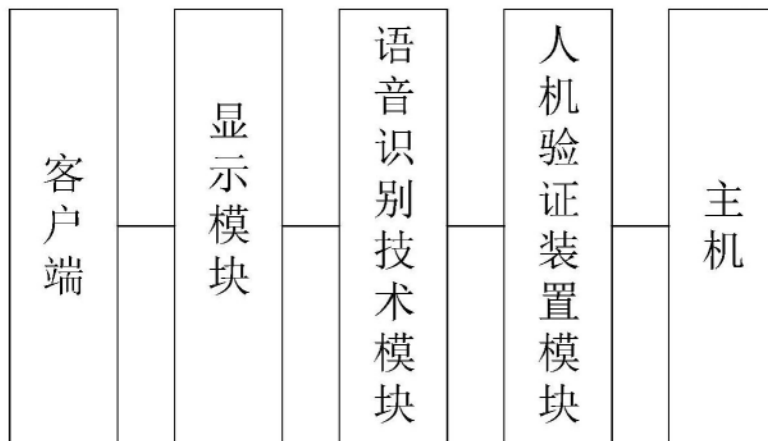


图1

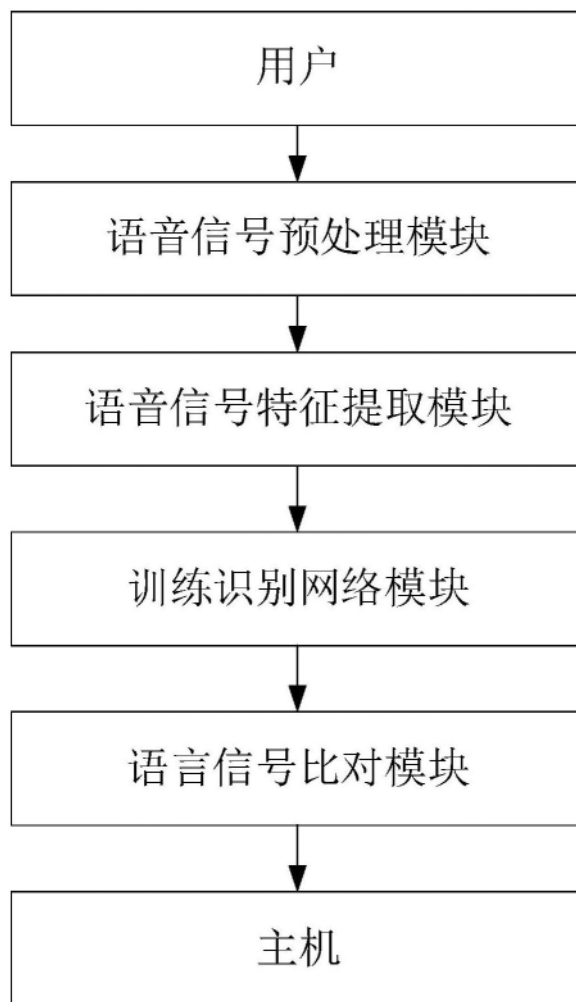


图2

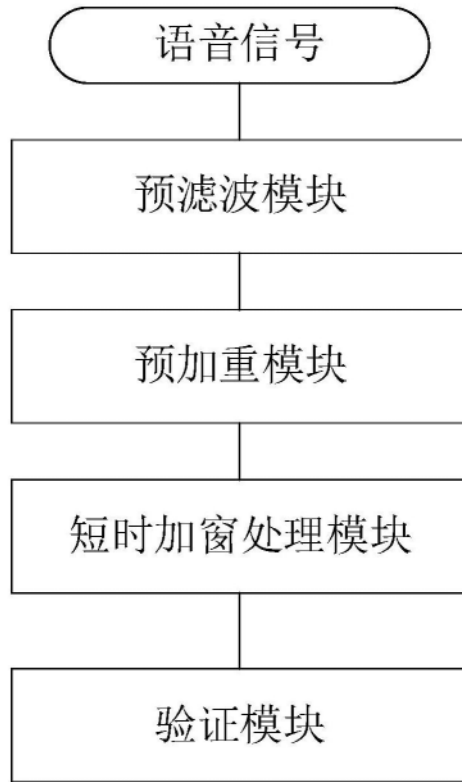


图3

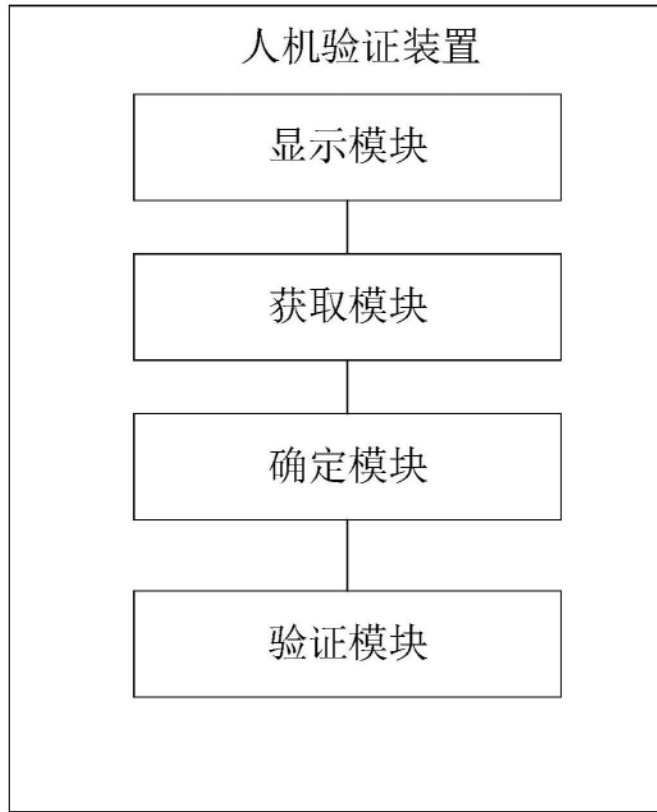


图4